REMARKS

Applicants request reconsideration and allowance of the subject application in view of the foregoing amendments and the following remarks.

Claims 122-136, 138, 140-153, 155, 157-160, 162-168, and 170-174 are pending in the present application, with Claims 122-124, 141-143, 168, 170, and 172-174 being independent.

Claims 137, 139, 154, 156, 161, and 169 have been cancelled without prejudice to or disclaimer of the subject matter contained therein. Claims 122-131, 133, 138, 141-146, 147-150, 152-153, 155, 159-160, 162-163, 165, 167-168, and 170 have been amended. Claims 171-174 are newly presented. No new matter is believed to have been added.

An Information Disclosure Statement, which addresses the Examiner's comments regarding the Information Disclosure Statement filed January 14, 2004, is being concurrently filed herewith.

Applicants acknowledge that Claims 165, 167, and 169 are considered to contain allowable subject matter. Applicants would respectfully like to point out that the noted patentable features in the Examiner's statement of reasons for allowability are not recited in each of the allowable claims. For example, Claim 165 does not recite "a header storage area." Nevertheless, Applicants submit that Claims 165, 167, and 169 are allowable for the combinations of features recited therein.

The specification was objected to as not containing antecedent basis for Claim 128. Applicants have amended the specification in response. Since support for the amendment

may be found at least in original Claim 4, no new matter has been added. Reconsideration and withdrawal of the objection to the specification are requested.

Claims 146 and 163 were objected to as containing informalities. These claims have been amended in response. Applicants request reconsideration and withdrawal of the objection to these claims.

Claims 169, which was also objected to, has been cancelled herein without prejudice or disclaimer. The objection is therefore submitted to be moot. Further, Claim 161 has been cancelled in response to the Examiner's advisory directed thereto.

Claims 149, 150, 152, and 153 were rejected under 35 U.S.C. § 112, second paragraph, as being indefinite. Amendments have been made in response, and reconsideration and withdrawal of the rejection to these claims are requested.

Claim 168 has been rejected under 35 U.S.C. § 101 as being directed to non-statutory subject matter. Claim 168, as amended, more clearly recites statutory subject matter, namely, a computer readable medium storing computer program code. Accordingly, reconsideration and withdrawal of the rejection to Claim 168 are requested.

Claims 122, 123, 131-133, 135-137, 141, 142, 148-150, 152-154, 168, and 170 have been rejected under 35 U.S.C. § 102(b) as being anticipated by U.S. Patent No. 5,638,425 ("Meador"). Claims 124, 138-140, 143, and 155-157 have been rejected under 35 U.S.C. § 102(e) as being anticipated by U.S. Patent No. 6,172,675 ("Ahmad").

Claims 122, 123, 125-130, 134, 141, 144-147, 151, 158, 159, 162-164, and 166 have been rejected under 35 U.S.C. § 103(a) as being obvious over Ahmad in view of U.S. Patent No. 5,787,414 ("Miike"). Claims 160-161 have been rejected under 35 U.S.C. § 103(a) as being

obvious over Ahmad in view of Miike and the paper "A Formal Framework for Linguistic Annotation" ("Bird").

These rejections are respectfully traversed.

Information retrieval systems have been proposed in which text annotations are provided for data files in order to allow users to search for portions within the data files by finding the corresponding portions within the text. Ahmad and Miike, for example, describe such systems.

The problem with these systems is that they are text-based. Therefore, if the annotation or the subsequent query is generated using an automatic speech recognition system, it may not be possible to retrieve a desired file because of recognition errors made by the speech recognizer or because of words that are outside of a vocabulary. According to the invention, this problem can be overcome by generating annotation data which includes the combination of words and phonemes which will allow an associated data file to be retrieved by searching for words and phonemes (although not necessarily simultaneously) within the annotation data.

Applicants submit that none of the cited documents, whether taken alone or in combination, teaches or suggests annotation data that can include the combination of both words and phonemes which can allow users to retrieve a data file by searching for words and phonemes within the annotation data.

Meador discloses an automatic directory assistance system. Regarding the reasons given for the § 102(b) rejection over Meador, Applicants submit that they constitute a comparison of the claimed features of the invention with aspects of the retrieval system described

by Meador, and not the aspects of Meador that relate to the way in which annotation data is generated.

In particular, while Meador teaches that a spoken utterance can be converted into phonemes and words, this is done for retrieval purposes. Even if the town name and address information stored within a database according to Meador is taken to be "annotation data," there is no disclosure or suggestion of generating the stored city names using speech recognition.

With regard to the retrieval operation described by Meador, this operates by receiving a spoken query from the user. The spoken query is then processed by a word-based recognizer and a separate phoneme-based recognizer. These separate recognizers operate to convert the user's query into words and phonemes, respectively, together with a probability value representing the confidence that the recognizer has made a correct recognition. The system weights the word probabilities output by the word recognizer with the phoneme probabilities output by the phoneme recognizer and then makes a final decision on the spoken word based on the combined probability. The system then retrieves the information from the database based on the word (text) with the highest combined probability. There is no disclosure or suggestion of searching the database using phonemes.

Applicants submit that there is no disclosure or suggestion in Meador of a word decoder which processes phoneme data generated by an automatic speech recognizer to identify words within the phoneme data, as recited in independent Claim 122 (and as similarly or correspondingly recited in independent Claims 123, 141, 142, 168, and 170). As discussed above, Meador teaches the use of separate word recognizer and phoneme recognizer. Further, Meador does not teach or suggest the storage of annotation data (which includes the generated

phoneme data and the words identified by the word decoder) in a database to allow an associated data file to be retrieved by searching for words and phonemes within the stored annotation data.

Ahmad teaches the generation of text data for an audio/video file. In order to work as annotation data, the text must be time-aligned with the audio and/or video data. As acknowledged at column 11, lines 45-52, this alignment is achieved automatically if the text is generated using a speech recognizer. However, if the text is generated by some other means, then an alignment must be performed between the text and the audio data.

Ahmad achieves this alignment by using a word-to-phoneme dictionary to identify the sequences of phonemes representing each word in the text. The corresponding sequences of phonemes are then concatenated together to represent the text. Each phoneme within this concatenated phoneme string is then represented by an acoustic Hidden Markov Model (HMM). The audio data is then compared with the string of acoustic HMMs representing the text. This comparison results in an alignment between the audio data and the text. Once the alignment has been achieved, manipulations on the text can be translated into manipulations on the audio and/or video.

Independent Claims 124 and 143 have been amended to further include a storage device operable to store (or a step of storing) the annotation data in the database to allow the associated data file to be retrieved by searching for words and phonemes within the stored annotation data. Applicants submit that Ahmad does not teach or suggest the storage of combined word and phoneme annotation data in a database to allow an associated data file to be retrieved by searching for words and phonemes.

Although Ahmad implies in its introduction that its invention is suitable for searching for particular content in a video tape, there is no disclosure or suggestion of achieving this searching using phonemes. Indeed, the only embodiments disclosed are those where the manipulation is carried out on text. Therefore, a person of ordinary skill in the art, reading Ahmad as a whole, would interpret the searching referred to in the introduction as being the searching of video tape using the text.

Miike describes a data retrieval system which generates annotation data which can be subsequently searched to retrieve an associated data file. In one embodiment, Miike discloses that annotation data can be generated from audio signals using an automated speech recognizer which recognizes words in the audio. These words are then stored and can be used for subsequent retrieval operations.

As part of the process for identifying words within the audio, Miike mentions that the speech recognizer can generate phonemes and, from those phonemes, words within the audio. However, Miike does not teach or suggest using the phonemes for any other purpose. In particular, Applicants submit that Miike does not teach or suggest generating annotation data by combining the words with the phonemes and storing the combined phoneme and word annotation data for use in subsequent retrieval operations.

Therefore, Applicants submit that even if a person of ordinary skill in the art were to combine the teachings of Ahmad and Miike, he or she would still not arrive at the claimed invention. In particular, as acknowledged by the Examiner, Ahmad does not describe in detail the operation of the speech recognizer that it uses. Therefore, using the teaching of Miike, a person of ordinary skill in the art might have used a phoneme-based recognizer to generate

phonemes from the audio, and then generate text from the phonemes. However, neither Ahmad nor Miike discloses or suggests generating annotation data by combining the words with the phonemes and storing the annotation data in a database for use in subsequently retrieving an associated data file.

As discussed above, Ahmad only generates the phonetic transcription of the text when the text itself is not generated from a speech recognizer. This is because the phonemes are only generated in Ahmad in order to determine a time alignment between the text and the audio. When the text is generated automatically by an automatic speech recognition system, this alignment is achieved automatically, and Ahmad teaches that in this case, there is no need for the phonemes. Therefore, there would be no need to maintain the phoneme data especially as Miike does not teach or suggest maintaining the phoneme data for anything other than for generating words corresponding to spoken utterance.

Therefore, even if a person of ordinary skill in the art were to combine the teachings of Ahmad and Miike, the claimed invention would still not be achieved.

Accordingly, Applicants submit that the none of the cited art, whether taken singly or in the combinations suggested, teaches or suggests all of the claimed features of the present invention. Withdrawal of the §§ 102-103 rejections is respectfully requested.

New independent Claims 173 and 174 are based on a combination of the existing independent Claims 122 and 123 and the additional features of allowable Claim 165. Applicants therefore submit that these new independent claims are allowable at least for the reasons indicated by the Examiner with respect to Claim 165. Further, new independent Claim 172 is submitted to be patentable for its claimed combination of features.

Applicants submit that the present invention is patentably defined by the independent claims. The dependent claims are also submitted to be allowable, for the reasons given regarding their respective independent claims, as well as due to the additional features they recite.

For example, Claim 160 depends from Claim 159, which in turn depends from Claim 158. According to these claims, annotation data can define a phoneme and word lattice which is arranged in a time-ordered sequence of blocks which is time-synchronized with a time-sequential signal. Further, each block of the phoneme and word lattice can include an associated time index identifying a timing of the block within the time-sequential signal. Claim 160 includes the feature that each node within the lattice represents a point in time at which a word and/or phoneme begins or ends within the associated time-sequential signal, and that each node includes a time offset value defining this point in time relative to the time index associated with the block. This is not taught or suggested by Ahmad, Miike, or Bird.

In particular, while Ahmad describes, during the alignment of the text with the audio, dividing the sequences of HMMs representing the text into blocks and then separately time-aligning each block with the audio, this division of the HMMs cannot be compared with the claimed annotation data as it is only an intermediate data structure which is never stored or intended to be stored in a database for use in subsequently retrieving the associated data file. The division into blocks described by Ahmad is only intended to reduce the processing burden required by the Viterbi time-alignment technique which is used.

With regard to Bird, Applicants submit that it does not teach or suggest that the nodes should have a time value representing the point of time within the time-sequential signal

that is defined relative to the time index of the block. All of the time values that are shown in Bird are defined in terms of the audio sample number. In other words, referring to section 3.1 of Bird, the word "she" is represented by the audio between samples 2360 and 5200. The word "had" is then represented by the audio between samples 5200 to 9680, and so on. Further, as shown in the next table in section 3.1, each of the phonemes forming the word "she" and "had" is defined with reference to the unique audio sample numbers and not relative to the time index of a block.

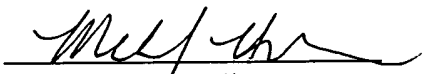
The Examiner also refers to section 4.8 in Bird where it is described that the annotations include time references and that "time function files" may have arbitrary offsets of their own. However, all of this is saying is that the stored audio file might have its own time line defined in its own terms (e.g., the time defined with reference to when the file was made or with reference to the time at which it was broadcast) and that the annotation data has its own time line, for example, starting at time t=0. However, this is not the same as dividing the phoneme and word annotation lattice into blocks, providing a time index for each block, and then referencing the time for each node within the block to the time index for the block.

Individual consideration of the dependent claims is respectfully solicited.

Favorable reconsideration and early passage to issue are respectfully requested.

Applicants' undersigned attorney may be reached in Washington, D.C. by telephone at (202) 530-1010. All correspondence should continue to be directed to the address given below.

Respectfully submitted,



Attorney for Applicants
Melody H. Wu
Registration No. 52,376

FITZPATRICK, CELLA, HARPER & SCINTO
30 Rockefeller Plaza
New York, New York 10112-3801
Facsimile: (212) 218-2200
MHW:ayr
171022 v 1